

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 19:43:59 ; Search time 9.25714 Seconds  
(without alignments)  
2096.859 Million cell updates/sec

Title: US-09-497-967-7

Perfect score: 2540

Sequence: 1 MKNILVILLISFINQIKS.....QCDFANFLSISLLISYLL 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	236.5	9.3	667	TS11_GIALA	Q03185 giardia lam
2	228.5	9.0	713	TS44_GIALA	P21849 giardia lam
3	189	7.4	328	CI170_GIALA	P15799 giardia lam
4	188	7.4	1609	LMG1_HUMAN	P11047 homo sapien
5	188	7.4	5376	ZAN_MOUSE	Q88799 mus musculo
6	185.5	7.3	1696	PK5_MOUSE	Q88799 mus musculo
7	184.5	7.3	687	VS41_GIALA	P92127 giardia lam
8	183.5	7.2	1680	FUR2_DROME	P30432 drosophila
9	179.5	7.1	1877	PK5_MOUSE	Q04592 mus musculo
10	179	7.0	2715	G156_PARPR	P13837 paramecium
11	178.5	7.0	1607	LMG1_MOUSE	P02468 mus musculo
12	175.5	6.9	3075	LM1_HUMAN	P25391 homo sapien
13	174.5	6.9	3718	LM4_MOUSE	Q61001 mus musculo
14	173	6.8	3712	LM4_DROME	Q00174 drosophila
15	167.5	6.6	2704	G168_PARPR	P17053 paramecium
16	166.5	6.6	969	PAC4_HUMAN	P29122 homo sapien
17	166	6.5	3084	LM1_MOUSE	P19137 mus musculo
18	165	6.5	1700	BAR3_CHITE	Q03376 chironomus
19	162	6.4	1576	YLK3_CAEEL	P41951 caenorhabdi
20	158	6.2	3106	LM2_MOUSE	Q60675 mus musculo
21	157	6.2	3110	LM2_HUMAN	P24043 homo sapien
22	155.5	6.1	1246	YVY2_CAEEL	P34504 caenorhabdi
23	155.5	6.1	3672	LM2_CAEEL	Q21313 caenorhabdi
24	154	6.1	1581	LMG3_MOUSE	Q97066 mus musculo
25	153.5	6.0	1790	LM1_DROME	P11046 drosophila
26	150.5	5.9	670	VG50_HSVII	Q00130 ictaluriid h
27	150.5	5.9	1169	YK82_YEAST	P36170 saccharomyc
28	150	5.9	1168	LMB3_MOUSE	Q61087 mus musculo
29	148.5	5.8	1639	LMG1_DROME	P15215 drosophila
30	148	5.8	3695	LM4_HUMAN	O15230 homo sapien
31	147	5.8	712	FBL1_CAEEL	O77469 caenorhabdi
32	146.5	5.8	937	PAC4_RAT	Q63415 rattus norv
33	146.5	5.8	1193	LMG2_HUMAN	Q13753 homo sapien

34	145.5	5.7	600	1	SP96_DICDI	P14328 dictyostell
35	145.5	5.7	790	1	ANP_NOTCO	P24856 notothenia
36	143.5	5.6	2471	1	NTC2_RAT	Q9q30 rattus norv
37	142	5.6	1167	1	XMRK_XIPMA	P13388 xiphophorus
38	142	5.6	1587	1	LMG3_HUMAN	Q9y66 homo sapien
39	141.5	5.6	610	1	LEM2_HUMAN	P16581 homo sapien
40	141.5	5.6	1786	1	LMB1_MOUSE	P02469 mus musculo
41	141	5.6	2911	1	YBN2_HUMAN	P35556 homo sapien
42	140.5	5.5	1416	1	YN81_CAEEL	Q03610 caenorhabdi
43	140	5.5	1064	1	FBP1_STRPU	Q10079 strongyloce
44	139.5	5.5	677	1	SP87_DICDI	P54643 dictyostell
45	139	5.5	768	1	LEM3_MOUSE	Q01102 mus musculo

## ALIGNMENTS

RESULT 1  
TS11\_GIALA  
AC Q03185;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Major surface trophozoite antigen II precursor.  
GN TSPII.  
OS Giardia lamblia (Giardia intestinalis).  
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.  
RN NCBI\_TaxID=5741;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Isolate AD-1;  
RX MEDLINE=932411215; PubMed=8479449;  
RA By P.L., Khanna K., Manning P.A., Mayrhofer G.;  
RT "A gene encoding a 69-kilodalton major surface protein of Giardia  
intestinalis trophozoites.";  
RL Mol. Biochem. Parasitol. 58:247-258(1993).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE  
PLASMA MEMBRANE.  
CC -1- DOMAIN: CONTAINS 27 REPEATS OF A CXCK MOTIF.  
CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: M95814; AAA02687.1; -;  
CC PIR: A48579; A48579.  
CC HSP: P02468; IKLO.  
CC InterPro: IPR000561; EGF-like.  
CC InterPro: IPR002174; Furin-like.  
CC InterPro: IPR005127; Giardia\_VSP.  
CC Pfam: PF03302; VSP; 1.  
CC SMART: SM00181; EGF; 3.  
CC SMART; SM00261; FU; 5.  
CC Antigen; Repeat; Transmembrane; Signal.  
CC SIGNAL 1 17  
CC CHAIN 18 667 MAJOR SURFACE TROPHOZOITE ANTIGEN II.  
CC DOMAIN 18 633 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 634 662 POTENTIAL.  
CC DOMAIN 663 667 CYTOPLASMIC (POTENTIAL).  
CC FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SEQUENCE 667 AA; 68475 MW; 1DD9572703232B8D CRC64;

Query Match 9.3%; Score 236.5; DB 1; Length 667;  
Best Local Similarity 22.7%; Pred. No. 7.5e-10;  
Matches 128; Conservative 49; Mismatches 193; Indels 195; Gaps 30;

```
QY 7 VILLISLFINIKSANGPVGTETNTAGVDDLGTPANCVMCO-----KNFY 52
DB 1 MLLAIFYFVITLFAKTTQTCEAKCEWVGTEICTRCQTKGVPIDKCVDTANAN 60
QY 53 YNNA-----AFVPGASTCTP-----CP-QKKDAGAQPNPPATANLVTCNV 93
DB 61 CKNASGDGDANOVGKMSVFGNTLCITVSPDGVCSVAANEYFPPNADATHDSVWSCSE 120
QY 94 KCPAGTA-----IAGAT-----DYAAIITEVCNCRINFYNAPNFNAGASTC 137
DB 121 ETPIHLANKQYIGVAGATCSAPKAGEDNTPKAATCTKCAAGFL--HTP--SEGUSC 176
QY 138 -TACPNVRVGGALTAGNAATIVACNVACPTGTALDDGVTTDYVRSFTECVKCRNFYFN 196
DB 177 EETCEPFGFHTAETSKT-----CK-SCITGSS-----EAPNVKGIQDCLKC---MYN 222
QY 197 GNGNTPFPNPKSQCTPCPAIKPANVAQATLGNDAIT--ITACQNV-----CPDG----- 244
DB 223 EASGNT-----LTCEKCSAQKPSL-DKTSNCDCTGONCAFCSSSGDRCGDSGFILD 275
QY 245 -----TISAAGVNNVAQNTCTNCAPNFNNAPNFNPGNSTCLPCPANKDYGA 294
DB 276 GQNCVKSCKTENCNCAKTPRAANEVCTECI-----STHLTTSQCVQVQALGNYIA 329
QY 295 EATAGGAATLAKQCNIA-----CPDGTALASGATNYVILQTECLNCAANFYFDGNNFQAGS 350
DB 330 GTNADNKA-CKEIVANCKTND-----QGOCQTCNDGDFYKNGDACSPCH 374
QY 351 SRCKACAPANKVOGAVATAGGTATLIAQALCEBPAGTVL--TDGTTSTYKQ----- 398
DB 375 ESKCTCSA-----GTA--SDCT-ECPTGKALKYGNDDGTGKFCGEGCTTGOGSG 419
QY 399 -----AASECVC-----AANFYTTKOTDWDVAGIDTCTSC----- 428
DB 420 ACKTGLTIDGASCEQDTONYEPQNGICSTTARTVATCKNSNVANGI--CSSCTNGF 477
QY 429 -----NKLTSGAEAN 439
DB 478 LRWNGCYETTFKPGKSVCEGANAD 502

RESULT 2
TS4_GIALA STANDARD; PRT; 713 AA.
ID TSA4_GIALA AC P21849;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major surface-labeled trophozoite antigen 417 precursor.
GN TSA 417.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxId=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 30957 / WB;
RX MEDLINE=90280395; PubMed=2352929;
RA Gillin F.D., Hagblom P., Harwood J., Aley S.B., Reiner D.S.,
RA McCaffery M., So M., Guiney D.G.;
RT "Isolation and expression of the gene for a major surface protein of
RT Giardia lamblia."
RL Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).
RN [2]
RP SEQUENCE OF 480-620 FROM N.A.
RC STRAIN=AD-1;
RX MEDLINE=93314970; PubMed=8325510;
RA Ey P.L., Mayrhofer G.;
RT "Two genes encoding homologous 70-kDa surface proteins are present
RT within individual trophozoites of the binucleate protozoan parasite
RT Giardia intestinalis."
RL Gene 129:257-262(1993).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
PLASMA MEMBRANE.
```

```
CC -!- DOMAIN: CONTAINS 29 REPEATS OF THE CXXC MOTIF.
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M33641; AAA02688.1; -
CC EMBL; M97488; AAA02581.1; -
CC PIR; A35502; A35502.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR005127; Giardia_VSP.
CC Pfam; PF03302; VSP; 2.
CC SMART; SM00181; EGF; 1.
CC SMART; SM00001; EGF_like; 1.
CC SMART; SM00261; FU; 3.
CC Signal; Antigen; Glycoprotein; Transmembrane; Repeat.
CC SIGNAL 1 17
CC CHAIN 18 713 MAJOR SURFACE-LABELLED TROPHOZOITE
FT ANTIEN 417
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 680 708
FT DOMAIN 709 713 POTENTIAL.
FT CARBOHYD 289 289 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 676 676 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 582 582 A -> T (IN STRAIN ADELAIDE-1).
FT VARIANT 606 606 A -> S (IN STRAIN ADELAIDE-1).
FT SEQUENCE 713 AA; 72510 MW; 9AD7195843DE5601 CRC64;
QY 23 CPVGTETNTAGVDDLGTPANCVCQKN-FYYN-----NAAAFVPGASTCTP-----CPQ 71
DB 69 CP-----QHSAGK-----CTQCGNSFMFKDGYSGEGLPGLSHLSLSSDGDGVCTE 115
QY 72 KKDAGAOPN---PPATANLVTCQNVK---PAGTAAGAGATDYAAIITEVCNCRINFYEN 126
DB 116 -----AAGYFAPVGAAN-TEQSVIACGDTTGVITAGGNTYKG-IADCAEC-----S 161
QY 127 APNFNAGA-----STCTACPNVRVGGALTAGNAATIVACQN----- 162
DB 162 APDATAGAEAGKVATCTKGVSK-----YLDKNVVCVDRKAOCNSGSTNKFVAVDDSENGKC 217
QY 163 VACPT-----GTALDDGVTTDYVRSFTECVKCRNFYNGNNGNTP-----FNP-- 206
DB 218 VCSNDNLNGVANDTCSDYDEQSKKIKCTCTDNNLYLKTTSSEGTSCVQKQCKRGDFPKD 277
QY 207 ---GKSOCTPCPAIKP--ANVAQATL--GNDATITACQNVACPDG----- 244
DB 278 DSSAGNKLCPNDSTDTGIANCATCALVSGRGAALVTCs-ACTDGYKPSADKTTCEAVSN 336
QY 245 ---TISAAGVNNVAQNTCTNCAPNFNNAPNFNPGNSTCLPCPA--NKDYGAETAGG 300
DB 337 CKTPGCKACSEKGENEVCTDCDGSY-----LTP-TSQCIDSCAKIGNYIGA---TEG 386
QY 301 AATLAKQCNIA---CPDGTALASGATNYVILQTECLNCAANFYFDGNNFQAGSSRCKAC 356
DB 387 AKLKCKECTAANCKTCD-----QCCQACNDGDFYKNGDACSPCHESCKTC 432
QY 357 PANKVOGAVATAGGTATLIAQALECPAGTVL---TDGTTSTYKQ-----AASECVCKA 407
DB 433 SA-----GTA---SDCT-ECPTGKALRYGDDGTGCTGCGCTTGTGAGACKTCG 477
QY 408 ANFYTTKOTDWDVAGIDTCTSC 428
DB 478 LT-----IDGASYCSEC 489
```

Query Match 9.0%; Score 228.5; DB 1; Length 713;  
Best Local Similarity 24.4%; Pred. No. 3e-09;  
Matches 122; Conservative 34; Mismatches 170; Indels 175; Gaps 33;



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
CC  
DR EMBL; M55210; AAA59492.1; JOINED.  
DR EMBL; M55217; AAA59492.1; JOINED.  
DR EMBL; M55201; AAA59492.1; JOINED.  
DR EMBL; M55211; AAA59492.1; JOINED.  
DR EMBL; M55212; AAA59492.1; JOINED.  
DR EMBL; M55213; AAA59492.1; JOINED.  
DR EMBL; M55214; AAA59492.1; JOINED.  
DR EMBL; M55215; AAA59492.1; JOINED.  
DR EMBL; M55216; AAA59492.1; JOINED.  
DR EMBL; M55192; AAA59492.1; JOINED.  
DR EMBL; M55193; AAA59492.1; JOINED.  
DR EMBL; M55194; AAA59492.1; JOINED.  
DR EMBL; M55195; AAA59492.1; JOINED.  
DR EMBL; M55196; AAA59492.1; JOINED.  
DR EMBL; M55197; AAA59492.1; JOINED.  
DR EMBL; M55198; AAA59492.1; JOINED.  
DR EMBL; M55199; AAA59492.1; JOINED.  
DR EMBL; M55200; AAA59492.1; JOINED.  
DR EMBL; M55202; AAA59492.1; JOINED.  
DR EMBL; M55204; AAA59492.1; JOINED.  
DR EMBL; M55205; AAA59492.1; JOINED.  
DR EMBL; M55206; AAA59492.1; JOINED.  
DR EMBL; M55207; AAA59492.1; JOINED.  
DR EMBL; M55208; AAA59492.1; JOINED.  
DR EMBL; M55209; AAA59492.1; JOINED.  
DR EMBL; J03202; AAA59488.1; -.  
DR EMBL; M27654; AAA59489.1; -.  
DR EMBL; X13939; CAA32122.1; -.  
DR PTR; S13548; MMH032.  
DR HSSP; P02468; 1TLE.  
DR Genew; HGNC:6492; LAMC1.  
DR MIM; 150290; -.  
DR InterPro; IPR004089; Chmtaxis\_transd.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001886; LamNT.  
DR InterPro; IPR000034; Laminin\_B.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR Pfam; PF00052; laminin\_B\_1.  
DR Pfam; PF00053; laminin\_EGF; 10.  
DR Pfam; PF00055; laminin\_Nterm; 1.  
DR PRINTS; PR00011; EGFLAMININ.  
DR ProDom; PD002082; LamNT; 1.  
DR ProDom; PD003031; Laminin\_B; 1.  
DR SMART; SM00180; EGF\_Lam; 10.  
DR SMART; SM00281; LamB; 1.  
DR SMART; SM00136; LamNT; 1.  
DR PROSITE; PS00022; EGF\_1; 8.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 11.  
DR Glycoprotein; Basament membrane; Extracellular matrix; Coiled coil;  
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
FT SIGNAL 1 33  
FT CHAIN 34 1609 LAMININ GAMMA-1 CHAIN.  
FT DOMAIN 34 285 LAMININ N-TERMINAL (DOMAIN VI).  
FT DOMAIN 286 341 LAMININ EGF-LIKE 1.  
FT DOMAIN 342 397 LAMININ EGF-LIKE 2.  
FT DOMAIN 398 444 LAMININ EGF-LIKE 3.  
FT DOMAIN 445 494 LAMININ EGF-LIKE 4.  
FT DOMAIN 495 504 LAMININ EGF-LIKE 5 (N-TERMINAL).  
FT DOMAIN 505 689 LAMININ DOMAIN IV.  
FT DOMAIN 690 723 LAMININ EGF-LIKE 6.  
FT DOMAIN 724 772 LAMININ EGF-LIKE 7.  
FT DOMAIN 773 827 LAMININ EGF-LIKE 8.  
FT DOMAIN 828 883 LAMININ EGF-LIKE 9.  
FT DOMAIN 884 934 LAMININ EGF-LIKE 10.  
FT DOMAIN 935 982 LAMININ EGF-LIKE 11.  
FT DOMAIN 983 1030 DOMAIN II AND I.  
FT DOMAIN 1030 1609 COILED COIL (POTENTIAL).  
FT DOMAIN 1038 1609

FT DISULFID 286 295 BY SIMILARITY.  
FT DISULFID 288 305 BY SIMILARITY.  
FT DISULFID 307 316 BY SIMILARITY.  
FT DISULFID 319 339 BY SIMILARITY.  
FT DISULFID 342 351 BY SIMILARITY.  
FT DISULFID 344 367 BY SIMILARITY.  
FT DISULFID 370 379 BY SIMILARITY.  
FT DISULFID 382 395 BY SIMILARITY.  
FT DISULFID 398 410 BY SIMILARITY.  
FT DISULFID 400 416 BY SIMILARITY.  
FT DISULFID 418 427 BY SIMILARITY.  
FT DISULFID 430 442 BY SIMILARITY.  
FT DISULFID 445 456 BY SIMILARITY.  
FT DISULFID 465 474 BY SIMILARITY.  
FT DISULFID 477 492 BY SIMILARITY.  
FT DISULFID 724 733 BY SIMILARITY.  
FT DISULFID 726 740 BY SIMILARITY.  
FT DISULFID 742 751 BY SIMILARITY.  
FT DISULFID 754 770 BY SIMILARITY.  
FT DISULFID 773 781 BY SIMILARITY.  
FT DISULFID 775 792 BY SIMILARITY.  
FT DISULFID 795 804 BY SIMILARITY.  
FT DISULFID 807 825 BY SIMILARITY.  
FT DISULFID 828 842 BY SIMILARITY.  
FT DISULFID 830 849 BY SIMILARITY.  
FT DISULFID 852 861 BY SIMILARITY.  
FT DISULFID 864 881 BY SIMILARITY.  
FT DISULFID 886 905 BY SIMILARITY.  
FT DISULFID 907 916 BY SIMILARITY.  
FT DISULFID 919 932 BY SIMILARITY.  
FT DISULFID 935 947 BY SIMILARITY.  
FT DISULFID 937 954 BY SIMILARITY.  
FT DISULFID 956 965 BY SIMILARITY.  
FT DISULFID 968 980 BY SIMILARITY.  
FT DISULFID 983 995 BY SIMILARITY.  
FT DISULFID 985 1001 BY SIMILARITY.  
FT DISULFID 1003 1012 BY SIMILARITY.  
FT DISULFID 1015 1028 BY SIMILARITY.  
FT DISULFID 1031 1034 INTERCHAIN (PROBABLE).  
FT DISULFID 1034 1034 INTERCHAIN (PROBABLE).  
FT DISULFID 1600 1600 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 60 60 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 650 650 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1161 1161 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1175 1175 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1205 1205 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1223 1223 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1241 1241 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1380 1380 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1395 1395 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1439 1439 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 212 212 F -> I (IN REF. 2).  
SQ SEQUENCE 1609 AA; 177606 MM; B098F20FCD97293B CRC64;  
Query Match 7.4%; Score 188; DB 1; Length 1609;  
Best Local Similarity 23.9%; Pred. No. 5.4e-06;  
Matches 99; Conservative 28; Mismatches 165; Indels 122; Gaps 26;  
QY 75 AGAQNPPATAMLVQCNVKCPAGTAAGATDAAIITECVNCRINFNENAPNFNAGA 134  
Db 674 ASARPGGVPAWVESCT--CPVG---YGGQF-----CEMC-LSGYRRETPNLGP-Y 718  
QY 135 STCTACPVNRVGGALTAGNAATIAQOC--NVACP\*GTALDDGVTTDYVRSFTECVKCRIN 192  
Db 719 SPCVLCAACN---GHSTCDPFGVCNCRNTAGP-----HCEKCSDG 757  
QY 193 FYYNGNGNTPFNPGKSQCTPCP-----AIKP-----ANVAQTLGNDAITTAQCN 238



```
FT CARBOHYD 1908      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1933      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2028      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2111      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2142      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2332      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2533      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2575      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2692      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2812      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3052      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3065      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3144      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3172      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3288      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3292      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3782      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4005      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4136      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4243      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4254      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4335      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4376      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4586      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5136      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5252      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 5376 AA; 579908 MW; 0E44DB77DF2A2620 CRC64;

Query Match      7.4%; Score 188; DB 1; Length 5376;
Best Local Similarity 21.2%; Pred. No. 1.8e-05;
Matches 114; Conservative 52; Mismatches 184; Indels 188; Gaps 30;

QY 17 QLSKANCVPGETNAGQVDDLTGPANCVN-----CQKNFYNNAAAFVPGASTCTP-C-- 69
Db 3151 QCNQCQCPKTYCK-----DLKSSNCTNIPLOCPAHSRTNC-----LPSPCPCLD 3199

QY 70 POKKDAGAQPNPATANLVTCNV-----KCPAGTAIAGGATDYAAIITECVNCRINF 122
Db 3200 PEGLCGTSKVPSTCREGICQCPGLMHKNC-----VLRIFGCKNTQCAF 3247

QY 123 YNENAPFNAGASTCTACPVNRVGALTAGNAATIVAQC-NVACPTGTALDDGVTTDYVR 181
Db 3248 ISADKTWISRGCTQSTCTP-----AGAI-----HCRNFKCPSGT----- 3281

QY 182 SFTECVKRLNFYNGNNGNTPFN-----PKSQCTPC-----PAIKP 219
Db 3282 -----YCKNGDNGSSNCTEITLQCPNQSFTDCLPSCVPSCSNRCVETSPSV-P 3329

QY 220 ANVAQATLGNDAFI-----TACNVACPDGTISAAGVNNVAAQ-----NTE 260
Db 3330 SSCREGCLCNHGTFVSEDKCVPRQTQCGCKDARGAIIIPAG-KTWTSKGCTQSCACVEGNIQ 3388

QY 261 CTN--CAPNFYNNAPFNPGNSTC-----LPCPANKDYGAATA-----GGAATLA 305
Db 3389 CNFQCPPEY-----CKDNSEGSSTKITLQCPAHTQYTSCLPCLPCLDPEGLCKDIS 3445

QY 306 KCQNIACPDGTAIAGATNVILQTECLNCAANFYFDGNNFQAGSSRCKACPAKRVQAV 365
Db 3446 PKVPSTCKEGCVQSG---YVLNSDKCVLRA-----ECDCRDAQAL 3484

QY 366 ATAGGTATL-----TAQC-ALECPAGTVLTDTSTYKQAASECVK-----CAA 408
Db 3485 IPAGKTWTSFGCTQSCACMGVAGVQCSQCPGTYCKDN-----EDGNSNCAKITLQCPA 3539

QY 409 NYFTTKQTDWAGIDTC-TSC--NKKLTSGAANLPESAKKNIOCDFANFSLISILLI 463
Db 3540 HSLFTN-----CLPPCLPCLDPLGCKGASPKVPSTCKEGCICQSGYVLSNNCKLL 3591

RESULT 6
PCK5_BRACL
ID PCK5_BRACL STANDARD; PRT; 1696 AA.
AC Q9NJ15; Q9NJ16; Q9NJ14;
```

```
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE (Proprotein convertase PC6-like) (aPC6).
GN PC6.
OS Branchiostoma californiensis (California lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7738;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS A, B AND C).
RP MEDLINE=20175281; PubMed=10708868;
RX Oliva A.A. Jr., Chan S.J., Steiner D.F.;
RA "Evolution of the prohormone convertases: identification of a
RT homologue of PC6 in the protochordate amphioxus.";
RL Biochim. Biophys. Acta 1477:338-348(2000).
CC -!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
CC PROPEPTIDES BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA
CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS
CC -!- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.
CC ISOFORM B IS A TYPE I MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B (SHOWN HERE) AND C; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
CC RETICULUM.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF184615; AAF26300.1; -
DR EMBL; AF184616; AAF26301.1; -
DR EMBL; AF184617; AAF26302.1; -
DR HSSP; Q99405; 1MPT.
DR MEROPS; S08.UPE; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002884; P_domain.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF01483; P; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PRODOM; PD000717; P_domain; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00261; FU; 17.
DR PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW Cleavage on pair of basic residues; Repeat; Alternative splicing;
KW Transmembrane.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 110 POTENTIAL.
FT CHAIN 111 1696 PROPROTEIN CONVERTASE SUBTILISIN/KEXIN
FT TYPE 5.
FT DOMAIN 111 1618 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1619 1639 POTENTIAL.
FT DOMAIN 1640 1696 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 111 488 CATALYTIC.
FT DOMAIN 496 637 HOMO B.
FT DOMAIN 664 1649 CYS-RICH MOTIF (CRM) REGION.
FT SITE 110 111 CLEAVAGE (AUTO-) (BY SIMILARITY).
```

```

FT ACT_SITE 192 192 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 233 233 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 407 407 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 885 885 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1259 1323 DDTLLDRGECITSCGGEYMDREKKKACHTCKECSDSEY
FT DDTCTACNCGFLTLTDASSCAGCP -> AENQNASFCFFA
FT PREVSVALAELHGLRLSYLTDVPPQSNPPDVTGLGADRRL
FT TTATSAGRA (IN ISOFORM C).
FT VARSPLIC 1324 1696 CHPTCKECSDYDDTCTACNCGFLTLTDASSCAGCPGQFL
FT VARSPLIC 1288 1343 HGGDCDSCHREKTC -> IARCVDRRBSWCDLVLRFNFC
FT VRYFVKRCGCTGKLYMDRPNRRSSQFTQGRN (IN
FT ISOFORM A).
FT VARSPLIC 1344 1696 MISSING (IN ISOFORM A).
FT SEQUENCE 1696 AA; 188410 MW; 281CBEL1784257CBD CRC64;

Query Match 7.3%; Score 185.5; DB 1; Length 1696;
Best Local Similarity 20.4%; Pred. No. 8.6e-06;
Matches 101; Conservative 43; Mismatches 189; Indels 161; Gaps 25;

QY 18 IKSANCPVGTETAG-----QVDDLGPANVCNQKFNYYNNAAFVPGASTCT 67
DB 1048 VRTNCPSFTYDDQDRECRPHDNCACDGPNNQNCCKEGFYKT-----PDGCS--T 1100
QY 68 PCPO---KKDAGAGPNP-----PATANLV-----TQCNVYKCPAGTAIAGG 104
DB 1101 GCPNRYKDDTNECKPCDSSCFTCSGPASFCLSCADGDFLHSSCRSTCPAG--FTGN 1158
QY 105 ATDYAALITCVNCRINFYNENAPNPNAGASTCTACPVNRVGGALTAGNAATIVAQCINVA 164
DB 1159 AESHECVESC-----EQDYYSSETGRCDPCYN-----CRACDNDGCAEC 1201
QY 165 CPTGTALD-----DGVTTDYVRSFTECVKCLNLYNNGNTPFPNGKSOCTPC 214
DB 1202 APIYIVVDGRCEETCEGEVQDRDRDAE--LSRC-----PCHQSKCTC 1245
QY 215 PATKANVAQATLGNATITACNVACPDGTISAAGVNNWVAQNTEC--TNCAPNFY--NNN 272
DB 1246 SG--PSD-----TDCDSCKGDDTILDRG-----ECITSCGPGFYMDRR 1281
QY 273 APNFNPGNSTCLPCPANKDYGAETA---GGAATLAKGNTACPDGTGTATIAQCALECPAGTVLT 389
DB 1282 EKCKKACHTCKEC--SDEYDTCACNDGFLTLTDASSCEAGCPGPGFLHGG----- 1331
QY 330 TECLNCAANFYFDGNNFQAGSRKACAPANKVQGAATAGGTATLTAQCALECPAGTVLT 389
DB 1332 -DCDSC-----HRECKTCGPHDNCILSCPGSYLNDQOCSTHCPEGT-- 1374
QY 390 DGTTSYKQAAEF---CVKCAANFYTKQTDWVAGIDTCTSC-----NKKLTSGA 436
DB 1375 ---EETVDSGETVLQCLRHVNCKTCHG---EGEEDCECANDIKYKQDGRCVTECQ 1427
QY 437 EANLPESAKKNIQC 450
DB 1428 EGHYPLTNECQC 1441

RESULT 7
VS41_GIALA
ID VS41_GIALA STANDARD; PRT; 687 AA.
AC P92127;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Variant-specific surface protein VSP4A1 precursor (CRISP-90).
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O2-4A1;
```

```

RX MEDLINE=97321154; PubMed=9178264;
RA Papanastasiou P., Bruderer T., Li Y., Bomelli C., Koehler P.;
RT "Primary structure and biochemical properties of a variant-specific
RT surface protein of Giardia.";
RT Mol. Biochem. Parasitol. 86:13-27(1997).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=97233006; PubMed=9078242;
RA Papanastasiou P., McConville M.J., Ralton J., Koehler P.;
RT "The variant-specific surface protein of Giardia, VSP4A1, is a
RT glycosylated and palmitoylated protein.";
RT Biochem. J. 322:49-56(1997).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -1- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC
CC AT THE REDUCING TERMINUS.
CC -1- PTM: PALMITOYLATED.
CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z83743; CAB06038.1; -.
CC HSSP; O14763; IDOG.
CC GlycoSuiteDB; P92127; -.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR005127; Giardia_VSP.
CC Pfam; PF03302; VSP; 2.
CC SMART; SM00181; EGF; 2.
CC SMART; SM00001; EGF-like; 1.
CC SMART; SM00261; FU; 3.
CC
CC Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;
CC Signal.
CC SIGNAL 1 14 POTENTIAL.
CC CHAIN 15 687 VARIANT-SPECIFIC SURFACE PROTEIN VSP4A1.
CC DOMAIN 15 660 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 661 681 POTENTIAL.
CC DOMAIN 682 687 CYTOPLASMIC (POTENTIAL).
CC SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;

Query Match 7.3%; Score 184.5; DB 1; Length 687;
Best Local Similarity 21.1%; Pred. No. 4.1e-06;
Matches 119; Conservative 54; Mismatches 214; Indels 177; Gaps 29;

QY 12 SLFINKISANCPVGTET-----NTAGQVDDLGPANVCNQKFNYYNNAAFVPGASTCT 67
DB 124 SQIFQNKATPEKSECLCWDITDRNGVGV-ANCATC-----TAPASSTGPATCT 175
QY 68 PCP----QKDKAGAOPNPATANLVTCNVKCPAGTAGGATGATYAALITCVNCRINFY 123
DB 176 ECMAGTYKKSD-----TEC-AACHSDCATCSEAN-----NOCTSCETGY 215
QY 124 -----NENAPNFN-----AGASTCTACPVNRVGGALTAGNAATIVAQCNVACPTG---- 168
DB 216 LKSNOCVEKNTCNTNHYPDPTSMTCVACTVLDANCATCSFDSATAKGC-LTCNSNKIPR 274
QY 169 TALDDGVTTDYVRSFTECVKCLNLYNNGNTPFPNGKSOCTPCPAIKPANVAQNTLG 228
DB 275 TTL-DGTSTCVENYAGC-----QGADNELFMKEDQSACLLCGDTEKNSDKG-VA 323
QY 229 NDATITACQN-----VACPDGTISAAG-VNNWVAQN-----TECTNCAPNFYN 271
DB 324 NCRTCTKNANDSPPTCTACLDGYFLERGSCTTTCADNCATCSEATTEKCKICKAGFF-- 301
QY 272 NAPNFNPGNSTCLPCPANKDYG-----AEATAGGAATL-AKQC-----NIACPDG 315
DB 382 ---LASPGEGKICISDSTNNGGIDGCAECTKEPAGPLKCTCKPNRKPACTSDNYCTEK 438
```







Db 1379 -----LDGGLCMEC-LSSQYDTSATKCHDSCRCFGQFCKGCVPLHL 1427

QY 398 -QAASECVKRANFTYTKQTDWAGIDTCTSCN 429

Db 1428 DQLSQCVCSCNQTLAEKTSAA-----CCNCD 1456

RESULT 9

PK5\_MOUSE

AC Q04592; Q62040; STANDARD; PRT; 1877 AA.

DT 01-FEB-1995 (Rel. 31, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE (Protein convertase PC5) (Subtilisin/kexin type 5 precursor (EC 3.4.21.-))

DE (Convertase PC5) (PC6) (Subtilisin/kexin-like protease PC5)

DE (SPC6).

GN PC5K5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B).

RC STRAIN=ICR; TISSUE=Intestine;

RX MEDLINE=93327934; PubMed=8335106;

RA Nakagawa T., Murakami K., Nakayama K.;

RT "Identification of an isoform with an extremely large Cys-rich region

RT of PC6, a Kex2-like processing endoprotease.;"

RL FEBS Lett. 327:165-171(1993).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM PC5A).

RC TISSUE=Brain, and Intestine;

RX MEDLINE=93224489; PubMed=8468318;

RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K.,

RA Nakayama K.;

RT "Identification and functional expression of a new member of the

RT mammalian Kex2-like processing endoprotease family: its striking

RT structural similarity to PACE4.;"

RL J. Biochem. 113:132-135(1993).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM PC5A).

RC TISSUE=adrenal cortex;

RX MEDLINE=93342056; PubMed=8341687;

RA Lussan J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;

RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a

RT candidate protein convertase expressed in endocrine and

RT nonendocrine cells.;"

RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).

RN [4]

RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.

RX MEDLINE=97103178; PubMed=8947550;

RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K.,

RA Bendayan M., Seidah N.G.;

RT "The isoforms of protein convertase PC5 are sorted to different

RT subcellular compartments.;"

RL J. Cell Biol. 135:1261-1275(1996).

RN [5]

RP DEVELOPMENTAL EXPRESSION.

RX MEDLINE=96293359; PubMed=86598813;

RA Constam D.B., Calton M., Robertson E.J.;

RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone

RT morphogenetic proteins at distinct sites during embryogenesis.;"

RL J. Cell Biol. 134:181-191(1996).

RN [6]

RP DEVELOPMENTAL EXPRESSION.

RX MEDLINE=97436919; PubMed=9291583;

RA Rancourt S.L., Rancourt D.E.;

RT "Murine subtilisin-like proteinase SPC6 is expressed during embryonic

RT implantation, somitogenesis, and skeletal formation.;"

RL Dev. Genet. 21:75-81(1997).

RN [7]

RP FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY

CC

CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE

CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE

CC FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED

CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION

CC OF GROWTH FACTORS.

CC -1- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR

CC PROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA

CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.

CC -1- SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED

CC SECRETORY PATHWAY. PC5B IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO

CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH

CC EARLY ENDOSOMES.

CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; PC5B/LONG (SHOWN HERE)

CC AND PC5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: PC5A IS EXPRESSED IN MOST TISSUES BUT IS MOST

CC ABUNDANT IN THE INTESTINE AND ADRENALS. PC5B IS EXPRESSED IN THE

CC INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.

CC -1- DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO,

CC EXCEPT IN THE DEVELOPING NERVOUS SYSTEM, THE RIBS AND THE LIVER,

CC BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT

CC E6.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT

CC E7.5, INTENSE EXPRESSION IN EXTRAEMBRYONIC ENDODERM, AMNION AND

CC NASCENT MESODERM. AT 8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK

CC SAC FOLLOWED BY A CONFINATION TO DERMATOTOME COMPARTMENT. BETWEEN

CC E9.5 AND E11.5, ABUNDANT EXPRESSION IN AER (THICKENED ECTODERMAL

CC CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED

CC TO THE CONDENSING MESENCHYM SURROUNDING THE CARILAGE. AT THIS

CC STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTBRAL AND FACIAL

CC CARTILAGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5,

CC ABUNDANT EXPRESSION IN EPITHELIAL CELLS OF THE INTESTINAL VILLI.

CC ISOFORM A IS MOST ABUNDANT AT ALL STAGES BUT SIGNIFICANT LEVELS OF

CC ISOFORM B OCCUR AT E12.5.

CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE

CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC

CC RETICULUM.

CC -1- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN

CC SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS

CC WITH THE TGN SORTING PROTEIN PACS-1.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation.

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; D17583; BAA04507.1; -

CC EMBL; D12619; BAA02143.1; -

CC EMBL; L14932; AAY4636.1; -

CC PIR; JX0248; JX0248.

CC PIR; A48225; A48225.

CC HSSP; Q99405; 1MPT.

CC MEROPS; S08.076; -

CC MGD; MGI:97515; Pcsk5.

CC InterPro; IPR000561; EGF-like.

CC InterPro; IPR002174; Furin-like.

CC InterPro; IPR002884; P\_domain.

CC InterPro; IPR00209; Peptidase\_S8.

CC Pfam; PF00082; Peptidase\_S8; 1.

CC Pfam; PF01483; P\_PARTIAL.

CC PRINTS; PR00723; SUBTILISIN.

CC ProDom; PD000717; P\_domain; 1.

CC SMART; SM00181; EGF; 3.

CC SMART; SM00001; EGF-like; 2.

CC SMART; SM00261; FU\_22

CC PROSITE; PS00136; SUBTILASE\_ASP; 1.

CC PROSITE; PS00137; SUBTILASE\_HIS; 1.

CC PROSITE; PS00138; SUBTILASE\_SER; 1.

CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;

CC Cleavage on pair of basic residues; Repeat; Alternative splicing;

KW







RA Timpl R., Uitto J.;  
RT "Human laminin: cloning and sequence analysis of cDNAs encoding A, B1  
RT and B2 chains, and expression of the corresponding genes in human  
RT skin and cultured cells.";  
RL Lab. Invest. 60:772-782(1989).  
CC -|- FUNCTION: Binding to cells via a high affinity receptor, laminin  
CC is thought to mediate the attachment, migration, and organization  
CC of cells into tissues during embryonic development by interacting  
CC with other extracellular matrix components.  
CC -|- SUBUNIT: Laminin is a complex glycoprotein, consisting of three  
CC different polypeptide chains (alpha, beta, gamma), which are bound  
CC to each other by disulfide bonds into a cross-shaped molecule  
CC comprising one long and three short arms with globules at each  
CC end.  
CC THE ALPHA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ) AND  
CC LAMININ-3 (S-LAMININ).  
CC -|- SUBCELLULAR LOCATION: Extracellular.  
CC -|- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
CC COMPONENT).  
CC -|- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
CC -|- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.  
CC -|- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
CC -|- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.  
CC -|- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.  
CC -|- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X58531; CAA1418.1; -;  
DR PIR; S14458; S14458.  
DR HSP; Q60675; 1Q0U.  
DR Genew; HGNC:6481; LAM1.  
DR MIM; 150320; -;  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001886; LamNT.  
DR InterPro; IPR000034; Laminin\_B.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR InterPro; IPR001791; Laminin\_G.  
DR Pfam; PF00052; laminin\_B; 2; 15.  
DR Pfam; PF00053; laminin\_EGF; 15.  
DR Pfam; PF00054; laminin\_G; 5.  
DR Pfam; PF00055; laminin\_Nterm; 1.  
DR PRINTS; PR00011; EGF\_LAMININ.  
DR ProDom; PD002082; LamNT; 1.  
DR ProDom; PD003031; Laminin\_B; 2.  
DR SMART; SM00180; EGF\_Lam; 14.  
DR SMART; SM00001; EGF\_Like; 1.  
DR SMART; SM00281; LamB; 2.  
DR SMART; SM00282; LamG; 5.  
DR SMART; SM00136; LamNT; 1.  
DR PROSITE; PS00022; EGF\_1; 11.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 15.  
DR PROSITE; PS50025; LAM\_G\_DOMAIN; 5.  
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
FT Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 3075 LAMININ ALPHA-1 CHAIN.  
FT DOMAIN 18 269 LAMININ N-TERMINAL (DOMAIN VI).  
FT DOMAIN 270 326 LAMININ EGF-LIKE 1.  
FT DOMAIN 327 396 LAMININ EGF-LIKE 2.  
FT DOMAIN 397 453 LAMININ EGF-LIKE 3.  
FT DOMAIN 454 502 LAMININ EGF-LIKE 4.  
FT DOMAIN 503 512 LAMININ EGF-LIKE 5 (N-TERMINAL).  
FT DOMAIN 517 708 LAMININ DOMAIN IV 1 (DOMAIN IV B).  
FT DOMAIN 709 741 LAMININ EGF-LIKE 5 (C-TERMINAL).  
FT  
FT DOMAIN 742 790 LAMININ EGF-LIKE 6.  
FT DOMAIN 791 848 LAMININ EGF-LIKE 7.  
FT DOMAIN 849 901 LAMININ EGF-LIKE 8.  
FT DOMAIN 902 950 LAMININ EGF-LIKE 9.  
FT DOMAIN 951 997 LAMININ EGF-LIKE 10.  
FT DOMAIN 998 1043 LAMININ EGF-LIKE 11.  
FT DOMAIN 1044 1089 LAMININ EGF-LIKE 12.  
FT DOMAIN 1090 1149 LAMININ EGF-LIKE 13.  
FT DOMAIN 1150 1159 LAMININ EGF-LIKE 14 (N-TERMINAL).  
FT DOMAIN 1160 1361 LAMININ DOMAIN IV 2 (DOMAIN IV A).  
FT DOMAIN 1362 1401 LAMININ EGF-LIKE 15.  
FT DOMAIN 1403 1451 LAMININ EGF-LIKE 16.  
FT DOMAIN 1452 1508 LAMININ EGF-LIKE 17.  
FT DOMAIN 1509 1555 LAMININ EGF-LIKE 17.  
FT DOMAIN 1556 2116 DOMAIN II AND I.  
FT DOMAIN 2117 2297 LAMININ G-LIKE 1.  
FT DOMAIN 2305 2481 LAMININ G-LIKE 2.  
FT DOMAIN 2486 2673 LAMININ G-LIKE 3.  
FT DOMAIN 2713 2885 LAMININ G-LIKE 4.  
FT DOMAIN 2890 3070 LAMININ G-LIKE 5.  
FT DOMAIN 1706 1796 COILED COIL (POTENTIAL).  
FT DOMAIN 1968 1989 COILED COIL (POTENTIAL).  
FT DOMAIN 2088 2120 COILED COIL (POTENTIAL).  
FT SITE 2534 2536 CELL ATTACHMENT SITE.  
FT DISULFID 270 279 BY SIMILARITY.  
FT DISULFID 272 290 BY SIMILARITY.  
FT DISULFID 292 301 BY SIMILARITY.  
FT DISULFID 297 305 POTENTIAL.  
FT DISULFID 304 324 BY SIMILARITY.  
FT DISULFID 327 336 BY SIMILARITY.  
FT DISULFID 329 361 BY SIMILARITY.  
FT DISULFID 364 373 BY SIMILARITY.  
FT DISULFID 376 394 BY SIMILARITY.  
FT DISULFID 397 409 BY SIMILARITY.  
FT DISULFID 399 427 BY SIMILARITY.  
FT DISULFID 429 438 BY SIMILARITY.  
FT DISULFID 441 451 BY SIMILARITY.  
FT DISULFID 454 467 BY SIMILARITY.  
FT DISULFID 456 471 BY SIMILARITY.  
FT DISULFID 473 482 BY SIMILARITY.  
FT DISULFID 485 500 BY SIMILARITY.  
FT DISULFID 742 751 BY SIMILARITY.  
FT DISULFID 744 757 BY SIMILARITY.  
FT DISULFID 760 769 BY SIMILARITY.  
FT DISULFID 772 788 BY SIMILARITY.  
FT DISULFID 791 806 BY SIMILARITY.  
FT DISULFID 793 816 BY SIMILARITY.  
FT DISULFID 819 828 BY SIMILARITY.  
FT DISULFID 831 846 BY SIMILARITY.  
FT DISULFID 849 863 BY SIMILARITY.  
FT DISULFID 851 870 BY SIMILARITY.  
FT DISULFID 873 882 BY SIMILARITY.  
FT DISULFID 885 899 BY SIMILARITY.  
FT DISULFID 902 914 BY SIMILARITY.  
FT DISULFID 904 921 BY SIMILARITY.  
FT DISULFID 923 932 BY SIMILARITY.  
FT DISULFID 935 948 BY SIMILARITY.  
FT DISULFID 951 963 BY SIMILARITY.  
FT DISULFID 953 969 BY SIMILARITY.  
FT DISULFID 971 980 BY SIMILARITY.  
FT DISULFID 983 995 BY SIMILARITY.  
FT DISULFID 998 1007 BY SIMILARITY.  
FT DISULFID 1000 1014 BY SIMILARITY.  
FT DISULFID 1016 1025 BY SIMILARITY.  
FT DISULFID 1028 1041 BY SIMILARITY.  
FT DISULFID 1044 1056 BY SIMILARITY.  
FT DISULFID 1046 1063 BY SIMILARITY.  
FT DISULFID 1065 1074 BY SIMILARITY.  
FT DISULFID 1077 1087 BY SIMILARITY.  
FT DISULFID 1403 1412 BY SIMILARITY.  
FT DISULFID 1405 1419 BY SIMILARITY.  
FT DISULFID 1422 1431 BY SIMILARITY.  
FT DISULFID 1434 1449 BY SIMILARITY.

```
FT DISULFID 1452 1466 BY SIMILARITY.
FT DISULFID 1454 1476 BY SIMILARITY.
FT DISULFID 1479 1488 BY SIMILARITY.
FT DISULFID 1491 1506 BY SIMILARITY.
FT DISULFID 1509 1521 BY SIMILARITY.
FT DISULFID 1511 1528 BY SIMILARITY.
FT DISULFID 1530 1539 BY SIMILARITY.
FT DISULFID 1542 1553 BY SIMILARITY.
FT DISULFID 1556 1566 INTERCHAIN (PROBABLE).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 555 555 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 763 763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 952 952 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1407 1407 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1579 1579 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1596 1596 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1678 1678 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1689 1689 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1698 1698 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1717 1717 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 6.9%; Score 175.5; DB 1; Length 3075;
Best Local Similarity 21.2%; Pred. No. 8.1e-05;
Matches 110; Conservative 44; Mismatches 187; Indels 177; Gaps 30;

QY 18 IKSANCPVGTETNTAGQVDDLTGPANCVCKQNFYNNAAAFVPGASTCTPCPKOKDAGA 77
Db 706 VEHCECPQGY-----TGTSCBCLSGYRYVDGILF---GGICQPCCECHGHA-- 748
QY 78 QNPPTANLVTCQNVKCPAGTAAGGATDYAAIITECVNCRINFRYNNAPNFRNAGATC 137
Db 749 -----AECNVH---GVCI---ACAHNTTCVHCQCLPGYGPSPRGTFGDCQPC 791
QY 138 TACPVRVGGALTAGNAATIAQCNVACPTGTALDDG--VTTDYVR---SFTCEVCVKRLN 192
Db 792 -ACPL-----TIAS--NPFST-CHLNDGDEVVCDWCAPGYSANWCERCADG 834
QY 193 FYNNGNNGTNPNGKSOCTPCPAIKPANVAQATLGNDAITTAQCNVACPDGTTSAAGVN 252
Db 835 YYGN-----PTVPGES-CVPCDC--SGNVDPSEAGHCDSVTGEC-LKCLGNTDGA---- 880
QY 253 NWVAQNTCTNCAPFYNN--NAPN-----FNPNSTCLPCPANKD-- 291
Db 881 -----HCERCADGYGDVATKNCRACECHVKGSHSAVCHLETGLCDCKPNVTGQOCD 933
QY 292 -----YGAETAGGAATLAKQCNIACPDGTATIASGATNY-----VILQTECLNCAAN 338
Db 934 QCLHGYGLDSHG-----CRPCN--CSVAGSVSDGCTDEGQCHVCPVAGKRCDCRAHG 986
QY 339 FFDNNFQAQSSRCKAPANKVQAVATAGTATLIAQCALECPAGT-----VLTDGT 392
Db 987 FY-----AYQDGS--CTPCDCPHTQNTCDPETG-----ECV--CPPHQTGGKCECEDGH 1032
QY 393 TSTYKQAASECVKCAANYTTKQTDWAGIDTCTG-----AKKNI---QCD 451
Db 1033 WCYDEVCQACNCSLGVSTHRRCDVVTHGQCKSKFGRACDQCSGLGYRDFPCVPCDC 1092
QY 429 NKLTSGAEANLPES-----AKKNI---QCD 451
Db 1093 DLRGTSGDACNLEQGLCGCVETGACPKCNVFPQCN 1130

RESULT 13
LMA5_MOUSE STANDARD; PRT; 3718 AA.
AC Q61001; 09JH06;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
```

```
DE Laminin alpha-5 chain precursor.
GN LAMA5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-92 FROM N.A.
RA Timpl R., Sasaki T.;
RT "Completion of the N-terminal sequence of the murine Laminin alpha 5
chain.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 84-3718 FROM N.A.
RA STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RP MEDLINE=96081906; PubMed=7499364;
EX Miner J.H., Lewis R.M., Sanes J.R.;
RA "Molecular cloning of a novel laminin chain, alpha 5, and widespread
expression in adult mouse tissues.";
RL J. Biol. Chem. 270:28523-28526(1995).
RN [3]
RP REVISIONS.
RA Miner J.H., Lewis R.M., Sanes J.R.;
RP Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -1- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
CC ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).
CC -1- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
CC KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
CC AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 LAMININ DOMAINS IV.
CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ293593; CAB99255.1; -.
DR EMBL; U37501; AAC53430.1; -.
DR HSSP; P02468; ITLE.
DR MGD; MGI:105382; Lama5.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001886; LamNT.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00052; laminin_B; 1.
DR Pfam; PF00053; laminin_EGF; 19.
DR Pfam; PF00054; laminin_G; 2.
DR Pfam; PF00055; laminin_Nterm; 1.
DR ProDom; PD002082; LamNT; 1.
DR ProDom; PD003031; Laminin_B; 1.
DR PROSITE; PS00022; EGF_1; 19.
DR PROSITE; PS01186; EGF_2; 3.
```

```
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
DR PROSITE; ESS0025; LAM_G_DOMAIN; 5.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 40
FT CHAIN 41 3718 LAMININ ALPHA-5 CHAIN.
FT FT DOMAIN 41 304 LAMININ N-TERMINAL (DOMAIN VI).
FT FT DOMAIN 305 363 LAMININ EGF-LIKE 1.
FT FT DOMAIN 364 433 LAMININ EGF-LIKE 2.
FT FT DOMAIN 434 479 LAMININ EGF-LIKE 3.
FT FT DOMAIN 500 546 LAMININ EGF-LIKE 4.
FT FT DOMAIN 547 592 LAMININ EGF-LIKE 5.
FT FT DOMAIN 593 637 LAMININ EGF-LIKE 6.
FT FT DOMAIN 638 682 LAMININ EGF-LIKE 7.
FT FT DOMAIN 683 728 LAMININ EGF-LIKE 8.
FT FT DOMAIN 729 781 LAMININ EGF-LIKE 9.
FT FT DOMAIN 782 833 LAMININ EGF-LIKE 10.
FT FT DOMAIN 834 855 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT FT DOMAIN 856 1442 LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT FT DOMAIN 1443 1488 LAMININ EGF-LIKE 12.
FT FT DOMAIN 1489 1532 LAMININ EGF-LIKE 13.
FT FT DOMAIN 1533 1581 LAMININ EGF-LIKE 14.
FT FT DOMAIN 1582 1632 LAMININ EGF-LIKE 15.
FT FT DOMAIN 1633 1642 LAMININ EGF-LIKE 16 (N-TERMINAL).
FT FT DOMAIN 1643 1831 LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT FT DOMAIN 1832 1864 LAMININ EGF-LIKE 16 (C-TERMINAL).
FT FT DOMAIN 1865 1914 LAMININ EGF-LIKE 17.
FT FT DOMAIN 1915 1970 LAMININ EGF-LIKE 18.
FT FT DOMAIN 1971 2024 LAMININ EGF-LIKE 19.
FT FT DOMAIN 2025 2071 LAMININ EGF-LIKE 20.
FT FT DOMAIN 2072 2118 LAMININ EGF-LIKE 21.
FT FT DOMAIN 2119 2168 LAMININ EGF-LIKE 22.
FT FT DOMAIN 2169 2735 DOMAIN II AND I.
FT FT DOMAIN 2736 2933 LAMININ G-LIKE 1.
FT FT DOMAIN 2947 3119 LAMININ G-LIKE 2.
FT FT DOMAIN 3128 3296 LAMININ G-LIKE 3.
FT FT DOMAIN 3337 3511 LAMININ G-LIKE 4.
FT FT DOMAIN 3518 3689 LAMININ G-LIKE 5.
FT FT DOMAIN 2205 2257 COILED COIL (POTENTIAL).
FT FT DOMAIN 2330 2464 COILED COIL (POTENTIAL).
FT FT DOMAIN 2604 2621 COILED COIL (POTENTIAL).
FT FT DOMAIN 2639 2705 COILED COIL (POTENTIAL).
FT FT SITE 1723 1725 CELL ATTACHMENT SITE (POTENTIAL).
FT FT SITE 1839 1841 CELL ATTACHMENT SITE (POTENTIAL).
FT FT DISULFID 305 314 BY SIMILARITY.
FT FT DISULFID 307 327 BY SIMILARITY.
FT FT DISULFID 329 338 BY SIMILARITY.
FT FT DISULFID 341 361 BY SIMILARITY.
FT FT DISULFID 364 373 BY SIMILARITY.
FT FT DISULFID 366 398 BY SIMILARITY.
FT FT DISULFID 401 410 BY SIMILARITY.
FT FT DISULFID 413 431 BY SIMILARITY.
FT FT DISULFID 434 445 BY SIMILARITY.
FT FT DISULFID 436 452 BY SIMILARITY.
FT FT DISULFID 454 463 BY SIMILARITY.
FT FT DISULFID 466 476 BY SIMILARITY.
FT FT DISULFID 500 512 BY SIMILARITY.
FT FT DISULFID 502 521 BY SIMILARITY.
FT FT DISULFID 523 532 BY SIMILARITY.
FT FT DISULFID 535 544 BY SIMILARITY.
FT FT DISULFID 547 559 BY SIMILARITY.
FT FT DISULFID 549 566 BY SIMILARITY.
FT FT DISULFID 568 577 BY SIMILARITY.
FT FT DISULFID 580 590 BY SIMILARITY.
FT FT DISULFID 593 605 BY SIMILARITY.
FT FT DISULFID 595 611 BY SIMILARITY.
FT FT DISULFID 613 622 BY SIMILARITY.
FT FT DISULFID 625 635 BY SIMILARITY.
FT FT DISULFID 638 650 BY SIMILARITY.
FT FT DISULFID 640 656 BY SIMILARITY.
FT FT DISULFID 658 667 BY SIMILARITY.
FT FT DISULFID 670 680 BY SIMILARITY.
FT FT DISULFID 683 695 BY SIMILARITY.
FT DISULFID 685 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 716 726 BY SIMILARITY.
FT DISULFID 1443 1455 BY SIMILARITY.
FT DISULFID 1445 1462 BY SIMILARITY.
FT DISULFID 1464 1473 BY SIMILARITY.
FT DISULFID 1476 1486 BY SIMILARITY.
FT DISULFID 1533 1548 BY SIMILARITY.
FT DISULFID 1535 1555 BY SIMILARITY.
FT DISULFID 1557 1566 BY SIMILARITY.
FT DISULFID 1569 1579 BY SIMILARITY.
FT DISULFID 1582 1594 BY SIMILARITY.
FT DISULFID 1584 1601 BY SIMILARITY.
FT DISULFID 1603 1612 BY SIMILARITY.
FT DISULFID 1615 1630 BY SIMILARITY.
FT DISULFID 1665 1674 BY SIMILARITY.
FT DISULFID 1867 1881 BY SIMILARITY.
FT DISULFID 1884 1893 BY SIMILARITY.
FT DISULFID 1896 1912 BY SIMILARITY.
FT DISULFID 1915 1930 BY SIMILARITY.
FT DISULFID 1917 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1953 1968 BY SIMILARITY.
FT DISULFID 1971 1986 BY SIMILARITY.
FT DISULFID 1973 1993 BY SIMILARITY.
FT DISULFID 1996 2005 BY SIMILARITY.
FT DISULFID 2008 2022 BY SIMILARITY.
FT DISULFID 2072 2083 BY SIMILARITY.
FT DISULFID 2074 2090 BY SIMILARITY.
FT DISULFID 2092 2101 BY SIMILARITY.
FT DISULFID 2104 2116 BY SIMILARITY.
FT DISULFID 2119 2126 BY SIMILARITY.
FT DISULFID 2121 2133 BY SIMILARITY.
FT DISULFID 2135 2144 BY SIMILARITY.
FT DISULFID 2147 2166 BY SIMILARITY.
FT DISULFID 2169 2199 INTERCHAIN (PROBABLE).
FT DISULFID 2172 2172 INTERCHAIN (PROBABLE).
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 905 905 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1335 1335 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2021 2021 N-LINKED (GLCNAC. . .) (POTENTIAL).
Query Match 6.9%; Score 174.5; DB 1; Length 3718;
Best Local Similarity 20.8%; Pred. No. 0.00012;
Matches 93; Conservative 34; Mismatches 140; Indels 181; Gaps 25;
Qy 32 AGQVDDLGTAN-----CVNCQKNFYNNAAAFVPGASTCTPCPKQKDGAGQNPPTAN 86
Db 1826 ASNVELCMCPANYRGDSQCECAPGYRDTKGLFL--GRCVPC-----1865
Qy 87 LVTOC---NVKCPAGTAAGATDYAAITTECVNCRINFYNNAPNPNAGASTCTACPVN 143
Db 1866 ---QCHGHSRDLCPGSGICVGC-QHNTEGQDCRCRCPGVSSDDPSN---PASPCVSCP-- 1916
Qy 144 RVGGALTAGNATVAOCNVACPT-----GTALDDGYTTDYVR---SFTCEVCRLNFYY 195
Db 1917 -----CPLAVPSNNFADGCVLRNGRTQCLCRPGYAGASCERCAPGFFG 1959
Qy 196 NGNNGNTPEFNGKSOCTPCPAIKPANVAQATLGNDATITACQNV---ACPDGTISAAGVN 252
Db 1960 N-----PLVLG-SSCQPCDC-----SGNGDPNMFSDCDPLTGACRGLRHTTG-- 2002
Qy 253 NWVAQNTNECTNCAFNFYNNAPNPNNGNST---CLPCPANKDYGAETAGTAATLAKQC- 308
Db 2003 -----PHCERCAPGFYGN---ALLPGNCTRCDCSPC-----GTETCDPSG 2040
```



```

QY 309 NIACPDGTAIASGATNYVILQTECLNCAANFYFDGNFQAGSSRCKAC-----PANKVQ 362
DQ 2041 RCLCRAG: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2041 RCLCRAG: : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 363 GAVATAGGTATLTAQCALECPAGVTLDTGTTSTYKQARASECVKCAANFYTKQTDWVAGI 422
DB 2082 ECHPOSG-----QC--HCQPGT-----TGPOCLECAPGTW----- 2109
QY 423 DTCSCNKKLTSGABANLPESAKNNIQ 450
DB 2110 -----GLPEKGCRCQC 2121

RESULT 14
LMA_DROME STANDARD; PRT; 3712 AA.
AC 000174;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha chain precursor.
GN LANA OR LAMA.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93049203; PubMed=1425586;
RA Kusche-Gullberg M., Garrison K., Mackrell A.J., Fessler L.I.,
RA Fessler J.H.;
RT "Laminin A chain: expression during Drosophila development and
RT genomic sequence.";
RL EMBO J. 11:4519-4527(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=94038678; PubMed=8223265;
RA Henchcliffe C., Garcia-Alonso L., Tang J., Goodman C.S.;
RT "Genetic analysis of laminin A reveals diverse functions during
RT morphogenesis in Drosophila.";
RL Development 118:325-337(1993).
RN [3]
RP SEQUENCE OF 1762-3712 FROM N.A.
RX MEDLINE=92078147; PubMed=1744083;
RA Garrison K., Mackrell A.J., Fessler J.H.;
RT "Drosophila laminin A chain sequence, interspecies comparison, and
RT domain structure of a major carboxyl portion.";
RL J. Biol. Chem. 266:22899-22904(1991).
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -!- FUNCTION: DIVERSE FUNCTIONS DURING MORPHOGENESIS IN DROSOPHILA.
CC COMPLETE LOSS-OF-FUNCTION MUTATIONS LEAD TO LATE EMBRYONIC
CC LETHALITY. CERTAIN PARTIAL LOSS-OF-FUNCTION MUTATIONS GIVE RAISE
CC TO ESCAPER ADULTS, WHICH HAVE ROUGH EYES ASSOCIATED WITH CHANGES
CC IN CELL FATE AND PATTERN, MISSHAPEN LEGS AND DEFECTS IN WING
CC STRUCTURE.
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).
CC -!- TISSUE SPECIFICITY: NEWLY FORMED MESODERM AND LATER PROMINENTLY
CC EXPRESSED IN HEMOCYTES, WHICH ALSO SYNTHESIZE COLLAGEN IV.
CC -!- DEVELOPMENTAL STAGE: DURING MORPHOGENESIS, MOSTLY IN EMBRYO
CC DEVELOPMENT AT 10-12 HOURS.
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT

```

```

CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV (DOMAIN IV' IS NOT
CC SIMILAR TO LAMININ DOMAIN IV).
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M96388; AAA28662.1; -
CC EMBL; L07288; AAC37178.1; -
CC EMBL; M75882; AAA28661.1; -
CC HSP; P02468; 1TLE.
CC FlyBase; FBgn0002526; LANA.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR000034; Laminin_B.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR001230; Prenyl_site.
CC Pfam; PF00052; laminin_B_1.
CC Pfam; PF00053; laminin_EGF; 20.
CC Pfam; PF00054; laminin_G; 5.
CC Pfam; PF00055; laminin_Nterm; 1.
CC PRINTS; PR00011; EGFLAMININ.
CC ProDom; PD002082; LamNT; 1.
CC ProDom; PD003031; Laminin_B; 1.
CC SMART; SM00180; EGF_Lam; 17.
CC SMART; SM00001; EGF_Like; 1.
CC SMART; SM00281; LamB; 1.
CC SMART; SM00282; LamG; 5.
CC SMART; SM00136; LamNT; 1.
CC PROSITE; PS00022; EGF_1; 17.
CC PROSITE; PS01186; EGF_2; 5.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
CC PROSITE; PS00025; LAM_G_DOMAIN; 5.
CC KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC -----
CC SIGNAL 1 22
CC CHAIN 23 3712 LAMININ ALPHA CHAIN.
CC DOMAIN 25 272 LAMININ N-TERMINAL (DOMAIN VI).
CC DOMAIN 273 332 LAMININ EGF-LIKE 1.
CC DOMAIN 333 402 LAMININ EGF-LIKE 2.
CC DOMAIN 403 447 LAMININ EGF-LIKE 3.
CC DOMAIN 448 494 LAMININ EGF-LIKE 4.
CC DOMAIN 495 540 LAMININ EGF-LIKE 5.
CC DOMAIN 541 586 LAMININ EGF-LIKE 6.
CC DOMAIN 587 631 LAMININ EGF-LIKE 7.
CC DOMAIN 632 676 LAMININ EGF-LIKE 8.
CC DOMAIN 677 731 LAMININ EGF-LIKE 9.
CC DOMAIN 732 784 LAMININ EGF-LIKE 10.
CC DOMAIN 785 815 LAMININ EGF-LIKE 11 (INCOMPLETE).
CC DOMAIN 816 1374 DOMAIN IV'.
CC DOMAIN 1375 1420 LAMININ EGF-LIKE 12.
CC DOMAIN 1421 1465 LAMININ EGF-LIKE 13.
CC DOMAIN 1466 1513 LAMININ EGF-LIKE 14.
CC DOMAIN 1514 1564 LAMININ EGF-LIKE 15.
CC DOMAIN 1565 1574 LAMININ EGF-LIKE 16 (N-TERMINAL).
CC DOMAIN 1575 1775 LAMININ DOMAIN IV (DOMAIN IV').
CC DOMAIN 1776 1808 LAMININ EGF-LIKE 16 (C-TERMINAL).
CC DOMAIN 1809 1858 LAMININ EGF-LIKE 17.
CC DOMAIN 1859 1916 LAMININ EGF-LIKE 18.
CC DOMAIN 1917 1969 LAMININ EGF-LIKE 19.
CC DOMAIN 1970 2016 LAMININ EGF-LIKE 20.
CC DOMAIN 2017 2063 LAMININ EGF-LIKE 21.
CC DOMAIN 2064 2111 LAMININ EGF-LIKE 22.

```



CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: X52133; CAA36378.1; -.  
DR PIR: S09118.  
DR InterPro: IPR002895; Paramesium\_SA.  
DR Pfam: PF01508; Paramesium\_SA; 34.  
KW Signal; Repeat; Antigen; Membrane; GPI-anchor.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 2704 188G SURFACE PROTEIN.  
FT DOMAIN 106 2560 37 X 75 AA APPROXIMATE REPEATS.  
FT DOMAIN 1060 1424 VARIABLE DOMAIN, COMPRISES 4 ALMOST  
FT IDENTICAL REPEATS.  
SQ SEQUENCE 2704 AA; 278775 MW; 40EA0A0B18EE2119 CRC64;

Query Match 6.6%; Score 167.5; DB 1; Length 2704;  
Best Local Similarity 23.7%; Pred. No. 0.00027;  
Matches 129; Conservative 49; Mismatches 187; Indels 179; Gaps 33;

QY 18 IKSANCP--VGTE-TWTAGVDDLGTTPANCVN---COKNF-----YNNAAAFVPGASTC 66  
Db : : ||| ||| ||| ||| ||| ||| ||| : : |||  
1056 LRFANCASITGTLTICVTYDPCGVAN-VNGTACQEKLATCAAYLTONSCTSTAGTC 1114  
QY 67 TPCQKQKDAQNPPTANLVTC-----NVKCP---AGTAI----AG 103  
Db : : : : ||| : : ||| : : ||| : : |||  
1115 A-----WSGSACLTVDANVATECAITGTGLTNAICAGYNAKCTVNRAGTACQKKEAL 1168  
QY 104 GATDYAAIITEC--VNCRINFYNENAPNFNAGASTCTACP----- 141  
Db || ||| : : : : ||| : : ||| : : |||  
1169 CAT-YAAVQATCSQSDAGLCAMSGSACLTVDANVATECPYITGTGLTNAICAGYNAKCT 1227  
QY 142 VNRVGGALT-----AGNAATIVAQCNTVA--CP--TGTA 171  
Db ||| ||| : : ||| : : ||| : : ||| : : |||  
1228 VNRAGTACQKKEALCATYAAVQATCSQSDAGLCAMSGSACLTVDANVATECPYITGTGL 1287  
QY 172 DCGVTTDYVRSFTECVKRLNFYNGNNGNTFPNGKSQCTPCPAIKPANVAQATLGND 231  
Db : : ||| : : ||| : : ||| : : ||| : : |||  
1288 TDAICAGY-----NAKCTV-----NRAGTACQKKEALCATYAAVQ-ATCSQSDAGLCA 1334  
QY 232 TITAOC-----NVA--CPDGTISAAGVNNWV--AQNTCT-----NCAPNF 268  
Db : : ||| ||| ||| : : ||| : : ||| : : |||  
1335 WSGSACLTVDANVATECP--YITGTGLTNAICAGYNAKCTVNRAGTACQKKEALCA--- 1389  
QY 269 YNNAPNFNGNSTCLPCPANKDYGAETAGGAATIAKOCNACPDGTALASGATNYVI- 327  
Db : : ||| : : ||| : : ||| : : ||| : : |||  
1390 -----TYAAVQATC-----SQSDAGLCAMSGSACLTVDANVATECAVITGTGLTDAICA 1439  
QY 328 -LQTECLNCAANFYFDGNFQAGSRCKA-CPANKVQAVATAGGTATLIIAOCALCECPAG 385  
Db : : ||| : : ||| : : ||| : : ||| : : |||  
1440 GYNAKCTNLK-----DGTGCGDEKATCKLYTTQNK-----CTSQTGTGFL--SCLWFDNSC 1487  
QY 386 TVLTDGTTSTYKQAASECVKCA-----NFYTTKOTDMVAGIDTCTSCNK 430  
Db : : ||| : : ||| : : ||| : : ||| : : |||  
1488 SPITDVTCSAIVQSGLDHAQCOQAYSTGCTSVSDGSKCQDFKTT--CEQYAG--TALSCTK 1543  
QY 431 KLT 434  
Db ||  
1544 TATS 1547

Search completed: February 11, 2003, 19:47:08  
Job time : 23.2571 secs